Conjugate gradients (CG)

Gradient methods

Problem : Find x^* :: $Ax^* = b$

where $A \in \mathbb{R}^{n \times n}$, A symmetric positive definite (spd)

$b \in \mathbb{R}^n$ given

Solution : Find x^* as the minimizer of the functional $\phi : \mathbb{R}^n \to \mathbb{R}$

$$\phi(x) = \frac{1}{2}(Ax, x) - (b, x)$$

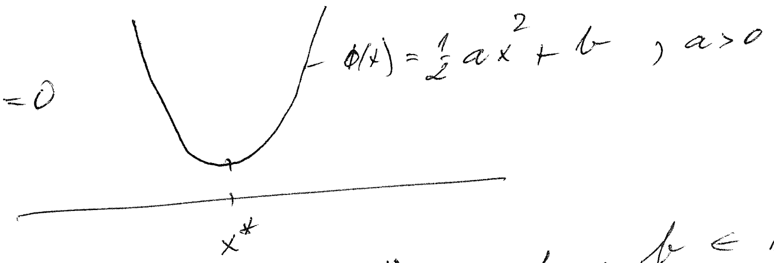
$(\cdot,\cdot)$ scalar product in $\mathbb{R}^n \times \mathbb{R}^n$

Motivation in 1D : $a > 0$, $b \in \mathbb{R}$ given

$ax^* = b \Leftrightarrow ax^* - b = 0 \Leftrightarrow x^*$ is the minimizer of $\phi(x) = \frac{1}{2}ax^2 - bx$,

since x^* is the minimizer of $\phi \Leftrightarrow \phi'(x^*) = 0$

$\phi'(x^*) = ax^* - b = 0$



Theorem : Let $A \in \mathbb{R}^{n \times n}$, A spd, $b \in \mathbb{R}^n$

It holds

$$Ax^* = b \Leftrightarrow \phi(x^*) < \phi(x) \quad \forall x \in \mathbb{R}^n,\ x \neq x^*$$

where $\phi(x) = \frac{1}{2}(Ax, x) - (b, x)$

Proof: $\boxed{\Leftarrow}$ If $x^* \in \mathbb{R}^n$ is the minimizer of ϕ

then $\nabla\phi(x^*) = 0$

Let us show that due to symmetry of A

$$\nabla\phi(x) = Ax - b$$

$$\frac{\partial \phi}{\partial x_k} = \frac{\partial}{\partial x_k}\left(\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_j x_i - \sum_{i=1}^{n} b_i x_i\right)$$

$$= \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}\left(\frac{\partial x_j}{\partial x_k} x_i + x_j \frac{\partial x_i}{\partial x_k}\right) - b_k$$

$$= \frac{1}{2}\left(\sum_{i=1}^{n} a_{i\cancel{j}k} x_i + \sum_{j=1}^{n} a_{kj} x_j\right) - b_k$$

$$= \frac{1}{2}\left(\sum_{i=1}^{n} a^T_{ki} x_i + \sum_{j=1}^{n} a_{kj} x_j\right) - b_k$$

$$\nabla\phi = \frac{1}{2}\left(A^T x + A x\right) - b$$

$$= Ax - b \quad \text{for } A = A^T \text{ (symmetry)}$$

$\boxed{\Rightarrow}$ due to the analogy with the Taylor expansion of quadratic function $f: \mathbb{R} \to \mathbb{R}$

$$f(x) = f(x^*) + f'(x^*)(x-x^*) + \frac{1}{2} f''(x^*)(x-x^*)^2$$

we have

$$\phi(x) = \phi(x^*) + (\underbrace{\nabla\phi(x^*)}_{Ax^* - b}, x - x^*) + \frac{1}{2}\left(A(x-x^*), (x-x^*)\right) \quad (*$$

This can be verified - the right hand side reads

$$\cancel{\frac{1}{2}(Ax^*, x^*)} - (b, x^*) + (Ax^*, x) - \cancel{(Ax^*, x^*)} - \underline{(b, x)} + (b, x^*)$$
$$+ \underline{\frac{1}{2}(Ax, x)} - \frac{1}{2}(Ax, x^*) - \frac{1}{2}(Ax^*, x) + \cancel{\frac{1}{2}(Ax^*, x^*)}$$

$$= \underline{\frac{1}{2}(Ax, x)} - \underline{(b, x)},$$

where we used the symmetry of A, i.e.

$$(Ax, x^*) = (x, A^T x^*) = (x, Ax^*) = (Ax^*, x)$$

From relation (*) we get

$$A x^* = b \implies \phi(x) > \phi(x^*) \qquad \forall x \neq x^*$$

as the consequence of the positive definiteness of A

<u>Finding the minimizer of ϕ</u>

at the step k, $x^{(k+1)}$ is computed as

$$x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}, \qquad k = 0, 1, 2, \dots,$$

where $p^{(k)}$ is the direction,

α_k is the length of the step along $p^{(k)}$,

with the goal

$$\phi(x^{(k+1)}) \leqq \phi(x^{(k)}), \quad \lim_{k \to \infty} x^{(k)} = x^*$$

<u>choice of α_k</u>

$x^{(k)}$ given, $p^{(k)}$ given, we seek α_k as the minimizer of $\phi(x^{(k)} + \alpha p^{(k)})$, $\alpha \in \mathbb{R}$

Differentiating with respect to α and setting it to zero yields the desired value of α_k

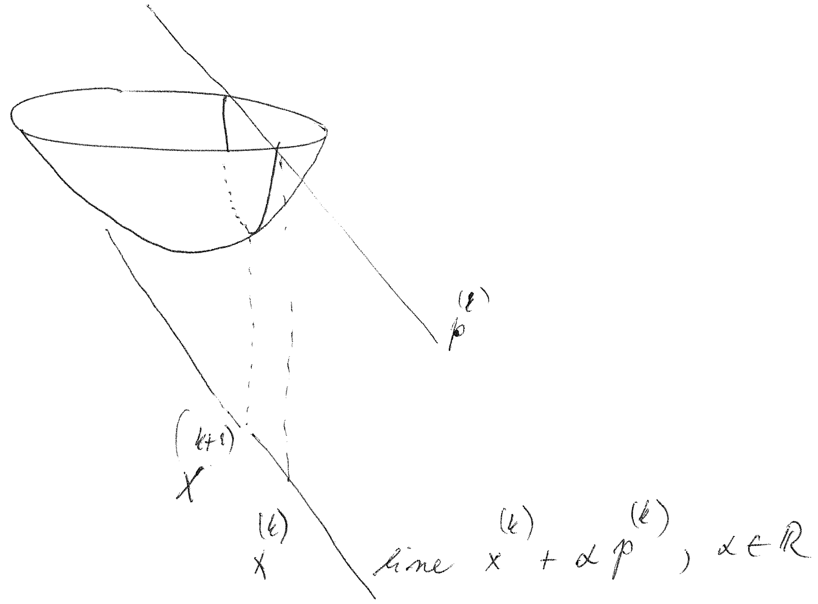
$$0 = \frac{d\,\phi(x^{(k)} + \alpha p^{(k)})}{d\alpha} = \left(\nabla\phi(x^{(k)} + \alpha p^{(k)}), p^{(k)}\right)$$

$$= \left(A(x^{(k)} + \alpha p^{(k)}) - b, p^{(k)}\right)$$

$$= \left(A x^{(k)} - b, p^{(k)}\right) + \alpha\left(A p^{(k)}, p^{(k)}\right)$$

$$\alpha_k := \frac{(r^{(k)}, p^{(k)})}{(A p^{(k)}, p^{(k)})}, \quad \text{where } r^{(k)} = b - A x^{(k)}$$

residual



Choice of $p^{(k)}$

$$x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}$$

We are interested in the residual $r^{(k+1)} = b - A x^{(k+1)}$

We have

$$\boxed{r^{(k+1)} = r^{(k)} - \alpha_k A p^{(k)}}$$

Let us note that from the definition of α_k we get

$$\left(r^{(k+1)}, p^{(k)}\right) = \left(r^{(k)}, p^{(k)}\right) - \underbrace{\frac{\left(r^{(k)}, p^{(k)}\right)}{\left(A p^{(k)}, p^{k}\right)}}_{\alpha_k} \left(A p^{k}, p^{(k)}\right) = 0$$

$$\boxed{\left(r^{(k+1)}, p^{(k)}\right) = 0 \quad , k = 0, 1, \dots}$$ by the definition of α_k

This leads us to the idea to choose $p^{(k)}$ in such a way that,

$$r^{(k+1)} \perp p^{(k)}, p^{(k-1)}, \dots, p^{(0)}$$

Then $r^{(n)} \perp \underbrace{p^{(n-1)}, p^{(n-2)}, \dots, p^{(0)}}_{n \text{ vectors}}$

and if $p^{(0)}, \dots, p^{(n-1)}$ are moreover LINEARLY INDEPENDENT, they form a basis of $\mathbb{R}^n$ and $r^{(n)} \perp \mathbb{R}^n$, namely $r^{(n)} \perp r^{(n)}$, so $r^{(n)} = 0$. Since $r^{(n)} = b - Ax^{(n)}$, we get the solution $x^* = x^{(n)}$.

<u>Sufficient conditions</u> imposed on $p^{(0)}, p^{(1)}, \dots, p^{(k)}$ such that

$$r^{(k+1)} \perp p^{(k)}, p^{(k-1)}, \dots, p^{(0)}$$

Let us remember

Method (M) $x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}$, $\alpha_k = \dfrac{(r^{(k)}, p^{(k)})}{(Ap^{(k)}, p^{(k)})}$

$r^{(k)} = b - Ax^{(k)}$

$x^{(0)}$ - arbitrary

$p^{(0)} := r^{(0)}$

Residual (R) $r^{(k+1)} = r^{(k)} - \alpha_k A p^{(k)}$

Property (P) $(r^{(k+1)}, p^{(k)}) = 0$
due to α_k

Wish (W) $r^{(k+1)} \perp p^{(k-1)}, p^{(k-2)}, \dots, p^{(0)}$

Result $x^{(n)} = x^*$ (solution of $Ax^* = b$)

Which are the sufficient conditions imposed on $p^{(0)}, p^{(1)}, \dots, p^{(k)}$ such that

$$\boxed{r^{(k+1)} \perp p^{(k)}, p^{(k-1)}, \dots, p^{(0)}}$$

We proceed in the following way. We set $p^{(0)} := r^{(0)}$

$\boxed{k=0}$ $\quad x^{(1)} = x^{(0)} + \alpha_0 p^{(0)} \Rightarrow (r^{(1)}, p^{(0)}) \overset{(P)}{=} 0$

$\boxed{k=1}$ $\quad x^{(2)} = x^{(1)} + \alpha_1 p^{(1)} \Rightarrow (r^{(2)}, p^{(1)}) \overset{(P)}{=} 0$ (for any $p^{(1)} \neq 0$)

wish $\quad (r^{(2)}, p^{(0)}) = 0$

$$(r^{(2)}, p^{(0)}) \overset{(R)}{=} (r^{(1)} - \alpha_1 A p^{(1)}, p^{(0)}) = \underbrace{(r^{(1)}, p^{(0)})}_{\overset{(P)}{=} 0} - \alpha_1 (A p^{(1)}, p^{(0)})$$

$$\boxed{(r^{(2)}, p^{(0)}) \overset{W}{=} 0 \Leftarrow (A p^{(1)}, p^{0}) = 0}$$

$\boxed{k=2}$ $\quad x^{(3)} = x^{(2)} + \alpha_2 p^{(2)} \Rightarrow (r^{(3)}, p^{(2)}) \overset{(P)}{=} 0$ (for any $p^{(2)} \neq 0$)

wish $\quad (r^{(3)}, p^{(1)}) = 0$

$$(r^{(3)}, p^{(1)}) \overset{(R)}{=} (r^{(2)} - \alpha_2 A p^{(2)}, p^{(1)}) = \underbrace{(r^{(2)}, p^{(1)})}_{\overset{(P)}{=} 0} - \alpha_2 (A p^{(2)}, p^{(1)})$$

$$\boxed{(r^{(3)}, p^{(1)}) \overset{W}{=} 0 \Leftarrow (A p^{(2)}, p^{(1)}) = 0}$$

wish $\quad (r^{(3)}, p^{(0)}) = 0$

$$(r^{(3)}, p^{(0)}) \overset{(R)}{=} \underbrace{(r^{(2)}, p^{(0)})}_{= 0 \ (k=1)} - \alpha_2 (A p^{(2)}, p^{(0)})$$

$$\boxed{(r^{(3)}, p^{(0)}) \overset{W}{=} 0 \Leftarrow (A p^{(2)}, p^{(0)}) = 0}$$

Conclusion (W) is satisfied, if $(A p^{(i)}, p^{(j)}) = 0$, $i \neq j$, $i, j = 0, \dots, k$
$(r^{(k+1)} \perp p^{(k)}, \dots, p^{(0)})$

i.e. $p^{(0)}, \dots, p^{(k)}$ are <u>A-orthogonal</u>

Note: A-orthogonal $\Rightarrow$ linearly independent

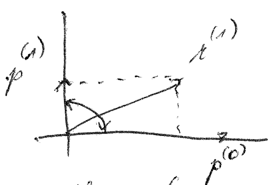
How to construct $p^{(0)}, p^{(1)}, \dots, p^{(n-1)}$ that are A-orthogonal?

We start with $x^{(0)}$ arbitrary, $p^{(0)} := r^{(0)} = b - Ax^{(0)}$

$$x^{(1)} = x^{(0)} + \alpha_0 p^{(0)} \quad (\text{see M for } \alpha_0)$$
$$r^{(1)} = r^{(0)} - \alpha_0 A p^{(0)}$$

we define $p^{(1)} := r^{(1)} - \beta_0 p^{(0)}$, where $\beta_0 : (Ap^{(1)}, p^{(0)}) = 0$

($r^{(1)}$ is A-orthogonalized against $p^{(0)}$)



A-orthogonal

$\boxed{\text{Then } p^{(0)}, p^{(1)} \text{ are A-orthogonal}}$

Further by induction over k

Let $p^{(0)}, \dots, p^{(k)}$ be A-orthogonal $(k \leq n-2)$

$x^{(1)}, \dots, x^{(k+1)}$ computed by the method (M)

$r^{(1)}, \dots, r^{(k+1)}$ residuals

we define $p^{(k+1)} = r^{(k+1)} - \beta_k p^{(k)}$, $\beta_k : (Ap^{(k+1)}, p^{(k)}) = 0$

($r^{(k+1)}$ is A-orthogonalized against $p^{(k)}$)

Then $(Ap^{(k+1)}, p^{(j)}) = 0$ $\boxed{j = 0, 1, \dots, k-1}$ so $p^{(0)}, \dots, p^{(k+1)}$ are A orthogonal

Proof

$$(Ap^{(k+1)}, p^{(j)}) = (p^{(k+1)}, Ap^{(j)}) = (\underbrace{r^{(k+1)} - \beta_k p^{(k)}, Ap^{(j)}}_{\text{A-orthog.}}) = (r^{(k+1)}, Ap^{(j)}).$$

Using $r^{(j+1)} = r^{(j)} - \alpha_j A p^{(j)}$, $j = 0, \dots, k-1$ we get

$$(r^{(k+1)}, Ap^{(j)}) = \Big(r^{(k+1)}, \underbrace{\frac{1}{\alpha_j}(r^{(j)} - r^{(j+1)})}_{Ap^{(j)}}\Big) = 0, \text{ since } r^{(k+1)} \perp r^{(0)}, r^{(1)}, \dots, r^{(k)}.$$

This is the consequence of the fact, that $r^{(k+1)} \perp \operatorname{span}\{p^{(0)}, \dots, p^{(k)}\}$ (in such a way $p^{(0)}, \dots, p^{(k)}$ were constructed)

and the fact that

$\mathrm{span}\{p^{(0)},\dots,p^{(k)}\} = \mathrm{span}\{r^{(0)},\dots,r^{(k)}\}$ due to

$$\boxed{C} \quad p^{(k+1)} = r^{(k+1)} - \beta_k p^{(k)}, \quad p^{(0)} = r^{(0)}$$

$$\boxed{D} \quad r^{(k+1)} = p^{(k+1)} + \beta_k p^{(k)}$$

so $\quad r^{(k+1)} \perp \mathrm{span}\{r^{(0)},\dots,r^{(k)}\}$

Determination of β_k

$$p^{(k+1)} = r^{(k+1)} - \beta_k p^{(k)} \quad \text{is} \quad (Ap^{(k+1)}, p^{(k)}) = 0$$

$$0 = (Ap^{(k+1)}, p^{(k)}) = (Ar^{(k+1)} - \beta_k Ap^{(k)}, p^{(k)}) \Rightarrow$$

$$\beta_k = \frac{(Ar^{(k+1)}, p^{(k)})}{(Ap^{(k)}, p^{(k)})}$$

Algorithm of conjugate gradient (CG) method

input $A, b, x^{(0)}$

$$r^{(0)} = b - Ax^{(0)}$$
$$p^{(0)} = r^{(0)}$$

for $k = 0, 1, \dots$

$$\alpha_k = \frac{(r^{(k)}, p^{(k)})}{(Ap^{(k)}, p^{(k)})}$$

$$x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}$$

$$r^{(k+1)} = r^{(k)} - \alpha_k Ap^{(k)}$$

$$\beta_k = \frac{(Ar^{(k+1)}, p^{(k)})}{(Ap^{(k)}, p^{(k)})}$$

$$p^{(k+1)} = r^{(k+1)} - \beta_k p^{(k)}$$

end

$\boxed{\text{Rounding errors}}$ CG is an iterative method (exercises)